Evaluating Attention Management Systems for Dynamic Monitoring Tasks

Anton Gasse
anton.gasse@uni-leipzig.de
ScaDS.AI, Leipzig University
Leipzig, Germany

Alexander Lingler
alexander.lingler@it-u.at
Interdisciplinary Transformation University
Linz, Austria

Martin Lorenz
lorenz@cs.uni-leipzig.de
ScaDS.AI, Leipzig University
Leipzig, Germany

Antti Oulasvirta
antti.oulasvirta@aalto.fi
Aalto University
Helsinki, Finland

Philipp Wintersberger
philipp.wintersberger@it-u.at
Interdisciplinary Transformation University
Linz, Austria

Patrick Ebel
ebel@uni-leipzig.de
ScaDS.AI, Leipzig University
Leipzig, Germany

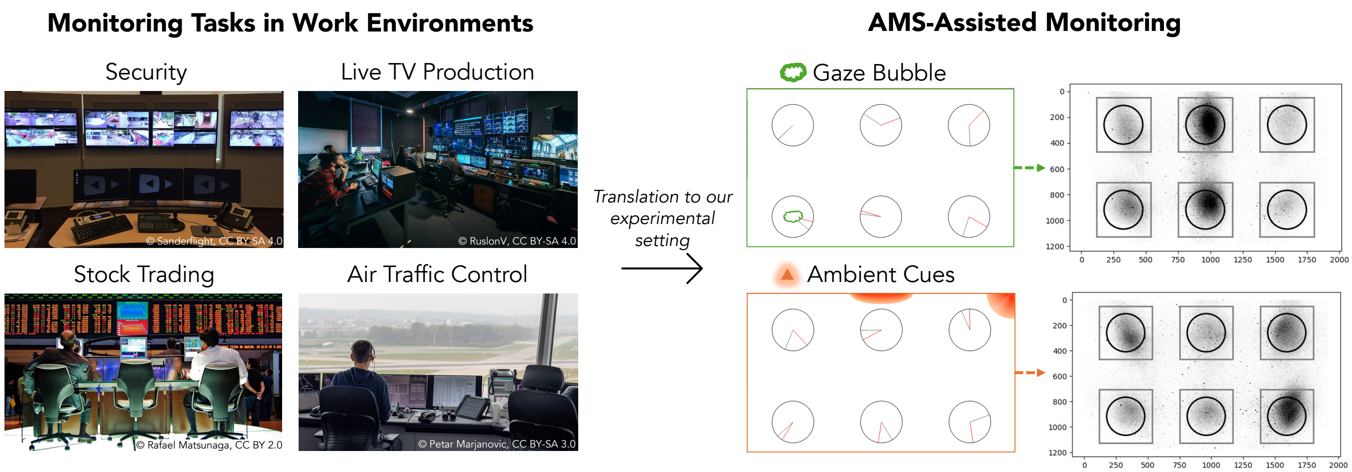


Figure 1: Left panel: Typical work environments that necessitate continuous, dynamic monitoring, wherein workers must observe various streams of information that differ in bandwidth. Right panel: Our experimental proxy to evaluate how AMS can assist in these tasks, building on Senders' Dial Task [28] and integrating two AMS designs aimed at enhancing monitoring performance: (1) a dynamically moving Gaze Bubble and (2) Ambient Cues. The heat maps visualize gaze distributions across the AMS designs and different dial configurations.

Abstract

In many work environments, operators must monitor multiple information sources, quickly identify critical situations, and respond appropriately. Attention Management Systems (AMS) are designed to help users coordinate attention in such contexts. However, while most AMS research has focused on multitasking and task-switching, their potential to guide gaze in dynamic monitoring remains unexplored. To address this, we evaluated two AMS designs in a controlled experiment (n=15) using Senders' Dial Task: Ambient Cues and a dynamic Gaze Bubble. Although participants were more likely to follow the moving Gaze Bubble, this design led to significantly poorer performance compared to Ambient Cues and a control group without AMS assistance. Our findings show that while AMS design influences visual attention, suboptimal designs can impair task performance. Further research is needed to identify design parameters that guide attention effectively while supporting performance.

CHIWORK '25 Adjunct, June 23–25, 2025, Amsterdam, Netherlands

ACM ISBN 979-8-4007-1397-2/25/06
https://doi.org/10.1145/3707640.3731920

CCS Concepts

• **Human-centered computing → User studies**; **Empirical studies in HCI**.

Keywords

Attention Management System, Situation Awareness, User Study

ACM Reference Format:
Anton Gasse, Alexander Lingler, Martin Lorenz, Antti Oulasvirta, Philipp Wintersberger, and Patrick Ebel. 2025. Evaluating Attention Management Systems for Dynamic Monitoring Tasks. In *CHIWORK '25 Adjunct: Adjunct Proceedings of the 4th Annual Symposium on Human-Computer Interaction for Work (CHIWORK '25 Adjunct), June 23–25, 2025, Amsterdam, Netherlands.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3707640.3731920

1 Introduction

During live TV production, the technical director must quickly decide which of multiple camera feeds to broadcast. In stock markets, traders must continuously monitor a wide range of stocks to make informed decisions. At airports, air traffic controllers must be aware of objects in the airspace, while security personnel ensure safety by inspecting numerous surveillance cameras. What all of these workplace situations have in common is that operators must monitor multiple sources of information of varying bandwidth, quickly identify potentially critical situations, and respond accordingly. To do so, operators must efficiently sample their environment and make decisions under uncertainty. However, the overwhelming flow of information can cause details to be overlooked [3]. Failure to maintain Situation Awareness (SA) [13, 30], even momentarily, can have serious consequences. Currently, high-level supervision becomes relevant in many workplace environments, given that routine tasks are increasingly handled by automation [29]. However, while automation can reduce workload, maintaining high levels of SA during monitoring tasks is inherently challenging, tiring, and can result in decreased task performance compared to non-automation settings [4, 23].

In this paper we study Attention Management Systems (AMSs), a class of interactive systems that has emerged as one potential solution to mitigate these *"ironies of automation"* [4], by offering context-sensitive support minimizing disruptions and optimizing information delivery [1, 3, 18]. Existing AMS manage high-level task switching and attention [1, 18] but do not focus on the low-level processes of visual sampling that are critical for SA in dynamic, high-stakes environments [8, 10, 22]. Guiding low-level attention is crucial when operators must rapidly shift from passive observation to active intervention, such as during automation failures. However, whether AMSs can enhance visual sampling and performance remains unclear. Prior research has focused on aggregated metrics (e.g., task performance, workload), leaving the link to visual sampling behavior underexplored [1]. Moreover, the design principles needed for AMSs to improve this aspect have received little attention. Thus, we investigate the following research questions:

RQ1: How does the design of AMSs influence monitoring performance in dynamic visual sampling tasks?
RQ2: Does the design of AMSs affect the rate at which participants' eye movements follow AMS suggestions?
RQ3: Does the design of the AMSs influence workload in monitoring tasks?

We contribute by exploring these research questions through a controlled experiment ($n = 15$) based on the Sender's Dial Task [12, 28], a well-established paradigm for studying dynamic monitoring tasks. In the Dial Task, users have to monitor six dials rotating at different speeds and respond by a button press when any of them is about to 'go critical'. We evaluate two radically different AMS designs inspired by previous work: (1) Ambient Cues that aim to guide users' attention via salient triggers in peripheral vision, and (2) a dynamic Gaze Bubble that guides attention by continuously moving between different targets. Our results inform the design of AMSs with the goal to improve visual sampling processes and task performance in monitoring tasks, ultimately benefiting many future workplace environments.

2 Background

Attention Management Systems. In general, AMSs aim to direct user's attention while interacting with computer-based systems in a way that reduces disruption and improves performance [1, 36]. While the underlying ideas are quite old [15, 34], successful implementations of the concept were, due to high computational complexity, demonstrated only recently [1, 18]. Lingler et al. [18] implemented an AMS to control interruption timing using reinforcement learning and computational rationality, and successfully demonstrated its ability to improve human performance in a fast-paced dual task. Yu et al. [37] presented a framework to derive better timings for suggestions in VR. Most works on AMS try to optimally distribute attention across multiple tasks — for example, by providing better interruption/notification timings or cues support the uptake of interrupted tasks [1, 31]. However, we argue that AMSs should be considered more widely and also support users within individual activities, such as low-level visual sampling processes. Here, an AMS could help users to sample the task environment as efficient and fast as possible.

Senders' Dial Task. The Dial Task by Senders [28] is a classic experimental paradigm for studying attention and visual sampling behavior in monitoring tasks. Participants observe a bank of dials, each with a different bandwidth, and must press a response key whenever a pointer crosses a threshold line from either direction (see the middle pane in Figure 1). This task requires participants to maintain a high level of SA while monitoring an environment with areas of varying stochasticity. This makes Senders' Dial Task an excellent proxy for real-world monitoring tasks as showcased in Figure 1. The Dial Task has recently been replicated with modern eye tracking experiment [11, 12].

3 AMS Designs

In this study, we investigate two AMS designs: Gaze Bubble and Ambient Cues, chosen to explore contrasting attention guidance strategies: one active and dynamic, and the other peripheral and ambient. The dynamic Gaze Bubble design draws on the human ability to track moving objects through smooth pursuit [26]. Inspired by the "butterfly guide" in Wallgrun et al. [33] and gaze sharing visualization techniques [2], it actively directs attention to specific regions. We designed the Gaze Bubble with a radius of 30 px and a green color. To make a suggestion, the Gaze Bubble moves between the centers of the different dials, always suggesting one dial at a time. The Ambient Cues design, inspired by Müller et al. [21], embodies the principles of calm interfaces [21, 34]. It uses ambient light displays to unobtrusively increase awareness without requiring a direct shift of attention. Ambient Cues to quickly guide attention have been researched in fields like driving or mixed and augmented reality [7, 20]. We implemented the Ambient Cues in a red-orange color gradient. The cues light up at fixed positions next to the dials they suggest. Unlike the Gaze Bubble, multiple Ambient Cues can be shown at once. The algorithm for triggering suggestions is largely similar for both AMSs. A dial is suggested when the pointer is within a 60 pixel distance on the x-axis and y-axis and is moving toward the threshold. If the algorithm suggests more than one dial, the Ambient Cues AMS highlights all relevant dials while the Gaze Bubble follows a pre-defined hierarchy.

4 Methodology

To assess how AMS design affects monitoring performance, eye movements, and workload depending on the difficulty of the monitoring task [12], we employed a within-subject design, manipulating both AMS design and task difficulty while collecting performance and eye tracking data.

Participants. We recruited 15 participants (3 female, 12 male) with a mean age of 24.8 years (SD=3.71) through an email invitation to participate in a user study. Each participant gave their written consent for participation and the collection of their data[1]. They received a 15€ compensation.

Experimental Design. We conducted a within-subject study employing a 3×2 factorial design to investigate the effects of AMSs design and task difficulty on monitoring performance, following rate and workload. The independent variables were the *AMS Design* with three levels: *No AMS* (baseline), *Ambient* (ambient peripheral cues) and *Gaze Bubble* (continuously moving indicator) and the *Difficulty* with two levels: *Easy* and *Hard*. The *Difficulty*, as originally formulated by Eisma et al. [12] is the amount of eye movement effort that is required to notice every threshold crossing.

We used the easiest and hardest configuration [12]. Monitoring performance is evaluated using a *Score*, awarding one point for each correct press and subtracting one point for each incorrect press. To evaluate how eye movements adjust to AMS suggestions, we compute the *Following Rate (%)*, representing the percentage of instances where participants followed AMS suggestions. Finally, we evaluate workload using pupil diameter as a proxy, following related work [6, 25, 35].

Apparatus. We closely followed the experimental setup of Eisma et al. [12] as outlined below [2]. Binocular eye movements were recorded at a sampling rate of 150 Hz using a GazePoint3 eye tracker[3]. The experiment was displayed on a DELL U2719D monitor with a resolution of 2560×1440 pixels and a display area of $596 \times 335\text{mm}^2$. The videos of the six dials were presented at 50 FPS with a resolution of 1920×1177 px. Each dial had a diameter of 316 px and was spaced 634 px horizontally and 658 px vertically. The Python script to generate the experimental videos is given in Sup. 3.

Procedure. During the experiment, participants were seated 95 cm away from the monitor, with the eye tracker 20 cm in front. A headrest ensured consistent positioning. Participants pressed the spacebar whenever a pointer in any dial crossed the threshold. They were informed that the AMSs would provide suggestions to improve monitoring. After calibrating the eye tracker, participants completed a 20 s familiarization trial for each AMS, followed by four monitoring tasks per condition (3 AMS, 2 difficulties), totaling 24 runs. After eight 90 s runs with one AMS, participants took a ≈ 1 min break. The conditions and threshold positions were randomized. Afterward, the participants completed a questionnaire on workload and AMS experience (see Sup. 5). Each experiment lasted ≈ 45 minutes.

5 Results

In the following sections, we address the research questions outlined in section 1. All analyses were conducted using R Statistical Software (v4.1.3) [24]. Due to multiple measurement per condition we used linear mixed effects models implemented with the *lme4* package (v1.1.34) [5]. We calculated p-values using Satterthwaite's degrees of freedom approximation [19], as implemented in the *lmerTest* package (v3.1.3) [16]. Post hoc pairwise comparisons were performed with the *emmeans* package (v1.8.7) [17] using Tukey's multiple comparison method [32]. Visual inspections of residual and Q-Q plots for the models revealed no notable deviations from homoscedasticity or normality. Regression tables were generated using the *stargazer* package (v5.2.3) [14]. Due to eye tracker calibration issues, one participant's data was excluded from models 3-6. All data can be found in Sup. 4.

RQ1: How does the design of AMSs influence monitoring performance in dynamic visual sampling tasks? To address RQ1, we first compare participants' scores across AMS conditions. To do this, we fit a linear mixed effects model with random intercepts (see model 1 in Table 1). We include *AMS* as a fixed effect and *participant* as a random effect. As also shown in Figure 2a, the results suggest that the AMS design affects monitoring performance. Post-hoc comparisons show that while there is no statistically significant difference between the *No AMS* and *AMS* conditions ($p = 0.710$), participants performed significantly worse, with a mean difference of ≈ 4 points, when supported by the Gaze Bubble AMS ($p < 0.001$ and $p = 0.002$). To test for an interaction effect between the AMS design and the difficulty of the monitoring task, we fitted a linear mixed effects model with difficulty as an additional fixed effect. The results (see Model 2 in Table 1) indicate that there is an interaction effect (see Figure 3 in the Appendix), and post hoc pairwise comparisons show that when participants were not assisted by an AMS, they performed significantly worse in the difficult condition ($p = 0.0197$). Given these results, we conclude that the design of AMS may negatively affect monitoring performance in tasks that require dynamic visual sampling.

RQ2: Does AMS design affect the rate at which participants' eye movements follow AMS suggestions? To investigate whether the AMS design affects the rate at which eye movements follow the AMS suggestions, we fit a linear mixed effects model with *AMS* as a fixed effect and *Participant* as a random effect. Our results (see Model 5 in Table 2 and compare Figure 2b) show that the rate at which eye movements follow Ambient Cues is significantly lower (−5.30%) compared to Gaze Bubble suggestions. To evaluate interaction effects with the task difficulty, we include *Difficulty* as an

[1]The consent form can be found in Sup. 1.
[2]An image of our setup is available in Sup. 2.
[3]https://www.gazept.com/product/gp3hd

Table 1: Linear Mixed Models for Score and Pupil Size

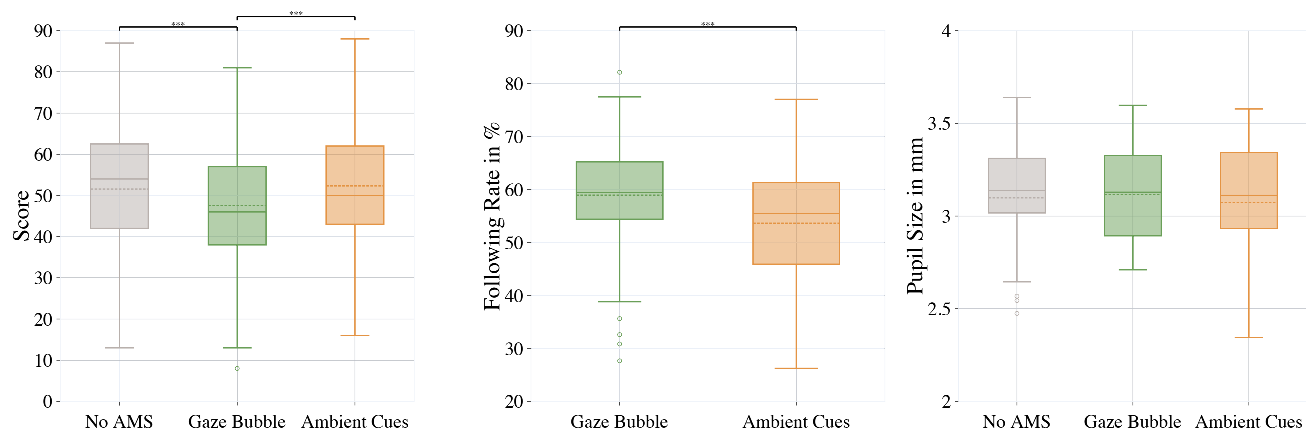
	Dependent variable:			
	Score		Pupil Size	
	Model 1	Model 2	Model 3	Model 4
Intercept	51.55*** (3.85)	53.15*** (3.91)	3.10*** (0.08)	3.10*** (0.08)
Ambient Cues	0.77 (0.97)	−0.32 (1.37)	−0.03 (0.03)	−0.04 (0.04)
Gaze Bubble	−3.97*** (0.97)	−4.95*** (1.37)	0.02 (0.03)	0.003 (0.04)
Hard		−3.20* (1.37)		−0.004 (0.04)
Ambient Cues : Hard		2.17 (1.93)		0.02 (0.05)
Gaze Bubble : Hard		1.95 (1.93)		0.03 (0.05)
AIC	2,541.49	2,533.23	−73.67	−54.20
BIC	2,560.92	2,564.32	−61.51	−34.76

*p<0.05; **p<0.01; ***p<0.001

Table 2: Linear Mixed Models for Following Rate

	Dependent variable:	
	Following Rate	
	Model 5	Model 6
Intercept	58.94*** (2.28)	62.78*** (2.36)
Ambient Cues	−5.30*** (0.97)	−7.93*** (1.26)
Hard		−7.67*** (1.26)
Ambient Cues : Hard		5.28** (1.78)
AIC	1,566.66	1,528.45
BIC	1,580.31	1,548.92

*p<0.05; **p<0.01; ***p<0.001



(a) Score **(b) Following Rate** **(c) Pupil Size**

Figure 2: Boxplots of the Score, Following Rate, and Pupil Size grouped by AMS Design. Statistically significant differences are indicated as: *p<0.05; **p<0.01; ***p<0.001

additional fixed effect (see Model 6 in Table 2). The results show a significant interaction effect and post-hoc pairwise comparisons indicate that the following rate of Gaze Bubble suggestions is $\approx 7.67\%$ ($p < 0.001$) lower for the hard dial configuration compared to the easy one. This effect does not apply to the Ambient Cues suggestions ($\beta \approx 2.40\%$, $p = 0.058$). Conclusively, our results indicate that AMS design affects the rate at which eye movements follow AMS suggestions. While participants are more likely to follow the Gaze Bubble AMS, the following rate decreases for harder tasks. This reduction is not significant for the Ambient Cues AMS.

RQ3: Do AMS suggestions influence workload in monitoring tasks? To investigate workload we compared participants pupil sizes across conditions. We fit a linear mixed effects model with *AMS* as fixed and *Participant* as random effect (Model 3 in Table 1) and another model that includes for interaction effects between *AMS* and *Difficulty* (Model 4 in Table 1). None of the models showed a significant difference in pupils size between conditions ($p > 0.05$ for all comparisons). Our post-hoc comparisons further revealed no significant difference in pupil size between the different difficulties. Therefore, based on our results, we have no indication that AMS design or AMS in general effect workload.

Qualitative Results. In the post-experiment interview, participants described their experience with the Ambient Cues AMS more positively than with the Gaze Bubble AMS. Six participants described the Gaze Bubble as being too fast, with an overlap of participants who also found it distracting and hard to ignore when false suggestions were made. One participant argued that the Gaze Bubble reduced his cognitive load. The Ambient Cues AMS was found to be helpful by seven participants. While three participants said that the many suggestions were distracting, three explicitly mentioned this feature as an advantage. In addition, the placement of the Ambient Cues in the periphery helped one participant to focus on other dials while already processing subsequent cues. In general, participants criticized the accuracy of the suggestions, indicating that fewer but more accurate suggestions would be appreciated.

6 Discussion

The main takeaway is that the design of an effective AMS interface is far from trivial. Our two AMSs, even if well-motivated and carefully implemented, had an impact on visual sampling strategies, yet they could not improve users' actual task performance. It remains an open question which, if any, AMS design can—effectively and efficiently—guide users to attend the right thing at the right time. In the remainder of this paper, we discuss two aspects of this challenge.

Performance Improvement May Require Better AMS Algorithms. Why did our two AMS designs fail to improve the participants' performance? We hypothesize that there are two interrelated causes: (1) the rule-based AMS implementation and (2) the user interface design. Given that participants followed the AMS suggestions fairly well, with following rates of more than 50 %, we do not believe their design alone is the main source of the performance issue.

Instead, our simple rule-based algorithm might be sophisticated enough to respond to the complex attentional dynamics involved in the task.

Succeeding in the monitoring task requires finding trade-offs between the user's current gaze position and their capability to quickly attend to and comprehend information in different locations.

Previous studies have demonstrated that AMSs may fare better if they adapt to the user's cognitive state and constraints [18]. Future research should explore whether a more advanced AMS algorithm utilizing model-based predictions alongside optimized visual cue designs can yield the desired performance benefits.

Performance Improvements May Cause Higher Workload. The following rates for the Gaze Bubble AMS were significantly higher than those for the Ambient Cues AMS, suggesting visual cue design influences users' visual sampling process.

Despite greater adherence to the Gaze Bubble AMS, task performance was significantly worse compared to the Ambient Cues AMS. This suggests that a poorly designed AMS can be detrimental, as users may follow its suggestions even when it negatively impacts their performance. These results underscore the importance of designing AMSs in which increased compliance is correlated with improved task performance, regardless of cue type. AMSs should be designed to guide user attention without generating confusion. Despite the differences in adherence to the visual cues, the overall workload was similar across conditions. If both the Gaze Bubble and Ambient Cues conditions required comparable levels of attention, this could explain why pupil size remained constant. However, in the baseline condition participants showed similar pupil dilation. It is possible that the demands of the monitoring task were already quite high, regardless of the support system. As we did not record the participants' pupil size in a resting condition, we cannot compare our results with other work due to the sensitivity to other factors such as individual differences or lighting conditions. However, since additional HMI elements such as the AMS suggestions typically increase the information processing load, the lack of differences in our experiment can be considered a positive result.

7 Conclusion

Our exploratory study represents a first step toward designing AMSs that guide low-level visual attention to assist workers in monitoring tasks. As workplace tasks become increasingly automated and the human role shifts to monitoring, effective attention management will be essential for maintaining high performance and worker well-being. A good AMS design should direct user attention to critical events without increasing workload or decreasing performance and user sense of control. This paper provides initial insights into how AMSs can support work in the future by highlighting caveats in the design of AMSs that can potentially lead to adverse effects. To improve openness and transparency in HCI research [9, 27], we make all research artifacts available on OSF: https://osf.io/kaers/.

Acknowledgments

The authors acknowledge the financial support by the Federal Ministry of Education and Research of Germany and by Sächsische Staatsministerium für Wissenschaft, Kultur und Tourismus in the programme Center of Excellence for AI-research „Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig“, project identification number: ScaDS.AI

References

[1] Christoph Anderson, Isabel Hübener, Ann-Kathrin Seipp, Sandra Ohly, Klaus David, and Veljko Pejovic. 2018. A Survey of Attention Management Systems in Ubiquitous Computing Environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (July 2018), 1–27. https://doi.org/10.1145/3214261

[2] Jad A. Atweh and Sara L. Riggs. 2024. Gaze Sharing, a Double-Edged Sword: Examining the Effect of Real-Time Gaze Sharing Visualizations on Team Performance and Situation Awareness. *Human Factors: The Journal of the Human Factors and Ergonomics Society* (Aug. 2024), 00187208241272060. https://doi.org/10.1177/00187208241272060

[3] Brian P. Bailey and Joseph A. Konstan. 2006. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in Human Behavior* 22, 4 (July 2006), 685–708. https://doi.org/10.1016/j.chb.2005.12.009

[4] L. Bainbridge. 1983. IRONIES OF AUTOMATION. In *Analysis, Design and Evaluation of Man–Machine Systems*. Elsevier, 129–135. https://doi.org/10.1016/B978-0-08-029348-6.50026-9

[5] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software* 67, 1 (2015), 1–48. https://doi.org/10.18637/jss.v067.i01

[6] Jackson Beatty and Brennis Lucero-Wagoner. 2000. The pupillary system. (2000). ISBN: 052162634X Publisher: Cambridge University Press.

[7] Leonardo Bonanni, Chia-Hsun Lee, and Ted Selker. 2005. Attention-based design of augmented reality interfaces. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*. ACM, Portland OR USA, 1228–1231. https://doi.org/10.1145/1056808.1056883

[8] J. C. F. De Winter, Y. B. Eisma, C. D. D. Cabrall, P. A. Hancock, and N. A. Stanton. 2019. Situation awareness based on eye movements in relation to the task environment. *Cognition, Technology & Work* 21, 1 (Feb. 2019), 99–111. https://doi.org/10.1007/s10111-018-0527-6

[9] Patrick Ebel, Pavlo Bazilinskyy, Mark Colley, Courtney Michael Goodridge, Philipp Hock, Christian P. Janssen, Hauke Sandhaus, Aravinda Ramakrishnan Srinivasan, and Philipp Wintersberger. 2024. Changing Lanes Toward Open Science: Openness and Transparency in Automotive User Research. In *Proceedings of the 16th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, Stanford CA USA, 94–105. https://doi.org/10.1145/3640792.3675730

[10] Patrick Ebel, Christoph Lingenfelder, and Andreas Vogelsang. 2023. On the forces of driver distraction: Explainable predictions for the visual demand of in-vehicle touchscreen interactions. *Accident Analysis & Prevention* 183 (April 2023), 106956. https://doi.org/10.1016/j.aap.2023.106956

[11] Yke Bauke Eisma, Ahmed Bakay, and Joost De Winter. 2024. Expectancy or Salience?—Replicating Senders' Dial-Monitoring Experiments With a Gaze-Contingent Window. *Human Factors: The Journal of the Human Factors and*

Ergonomics Society 66, 6 (June 2024), 1770–1785. https://doi.org/10.1177/00187208231176148

[12] Yke Bauke Eisma, Christopher D. D. Cabrall, and Joost C. F. De Winter. 2018. Visual Sampling Processes Revisited: Replicating and Extending Senders (1983) Using Modern Eye-Tracking Equipment. *IEEE Transactions on Human-Machine Systems* 48, 5 (Oct. 2018), 526–540. https://doi.org/10.1109/THMS.2018.2806200

[13] Mica R. Endsley. 1995. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37, 1 (March 1995), 32–64. https://doi.org/10.1518/001872095779049543

[14] Marek Hlavac. 2022. *stargazer: Well-formatted regression and summary statistics tables.* manual. Bratislava, Slovakia. https://CRAN.R-project.org/package=stargazer tex.organization: Social Policy Institute.

[15] Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems the CHI is the limit - CHI '99.* ACM Press, Pittsburgh, Pennsylvania, United States, 159–166. https://doi.org/10.1145/302979.303030

[16] Alexandra Kuznetsova, Per B. Brockhoff, and Rune H. B. Christensen. 2017. **lmerTest** Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82, 13 (2017), 1–26. https://doi.org/10.18637/jss.v082.i13

[17] Russell V. Lenth. 2022. Emmeans: Estimated marginal means, aka least-squares means.

[18] Alexander Lingler, Dinara Talypova, Jussi P. P. Jokinen, Antti Oulasvirta, and Philipp Wintersberger. 2024. Supporting Task Switching with Reinforcement Learning. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* ACM, Honolulu HI USA, 1–18. https://doi.org/10.1145/3613904.3642063

[19] Steven G. Luke. 2017. Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods* 49, 4 (Aug. 2017), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

[20] Andreas Löcken, Fei Yan, Wilko Heuten, and Susanne Boll. 2019. Investigating driver gaze behavior during lane changes using two visual cues: ambient light and focal icons. *Journal on Multimodal User Interfaces* 13, 2 (June 2019), 119–136. https://doi.org/10.1007/s12193-019-00299-7

[21] Heiko Müller, Jutta Fortmann, Martin Pielot, Tobias Hesselmann, Benjamin Poppinga, Wilko Heuten, Niels Henze, and Susanne Boll. 2012. Ambix: Designing ambient light information displays. In *Proceedings of Designing Interactive Lighting workshop at DIS*, Vol. 10. Citeseer. Issue: 2317956.2318081.

[22] Thanh Nguyen, Chee Peng Lim, Ngoc Duy Nguyen, Lee Gordon-Brown, and Saeid Nahavandi. 2019. A Review of Situation Awareness Assessment Approaches in Aviation Environments. *IEEE Systems Journal* 13, 3 (Sept. 2019), 3590–3603. https://doi.org/10.1109/JSYST.2019.2918283

[23] Raja Parasuraman and Victor Riley. 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 39, 2 (June 1997), 230–253. https://doi.org/10.1518/001872097778543886

[24] R Core Team. 2022. *R: A language and environment for statistical computing.* manual. Vienna, Austria. https://www.R-project.org/ tex.organization: R Foundation for Statistical Computing.

[25] Rahul Rajan, Ted Selker, and Ian Lane. 2016. Task Load Estimation and Mediation Using Psycho-physiological Measures. In *Proceedings of the 21st International Conference on Intelligent User Interfaces.* ACM, Sonoma California USA, 48–59. https://doi.org/10.1145/2856767.2856769

[26] D A Robinson. 1965. The mechanics of human smooth pursuit eye movement. *The Journal of Physiology* 180, 3 (Oct. 1965), 569–591. https://doi.org/10.1113/jphysiol.1965.sp007718

[27] Kavous Salehzadeh Niksirat, Lahari Goswami, Pooja S. B. Rao, James Tyler, Alessandro Silacci, Sadiq Aliyu, Annika Aebli, Chat Wacharamanotham, and Mauro Cherubini. 2023. Changes in Research Ethics, Openness, and Transparency in Empirical Studies between CHI 2017 and CHI 2022. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.* ACM, Hamburg Germany, 1–23. https://doi.org/10.1145/3544548.3580848

[28] John Warren Senders. 1983. *Visual Sampling Processes.* Doctoral Thesis. Tilburg University. https://pure.uvt.nl/ws/portalfiles/portal/1204640/3955241.pdf

[29] Thomas B. Sheridan and Raja Parasuraman. 2005. Human-Automation Interaction. *Reviews of Human Factors and Ergonomics* 1, 1 (June 2005), 89–129. https://doi.org/10.1518/155723405783703082

[30] N. A. Stanton, P. M. Salmon, G. H. Walker, E. Salas, and P. A. Hancock. 2017. State-of-science: situation awareness in individuals, teams and systems. *Ergonomics* 60, 4 (April 2017), 449–466. https://doi.org/10.1080/00140139.2017.1278796

[31] Dinara Talypova, Alexander Lingler, and Philipp Wintersberger. 2023. User-Centered Investigation of Features for Attention Management Systems in an Online Vignette Study. In *Proceedings of the 22nd International Conference on Mobile and Ubiquitous Multimedia.* ACM, Vienna Austria, 108–121. https://doi.org/10.1145/3626705.3627766

[32] John W. Tukey. 1949. Comparing Individual Means in the Analysis of Variance. *Biometrics. Journal of the International Biometric Society* 5, 2 (June 1949), 99. https://doi.org/10.2307/3001913

[33] Jan Oliver Wallgrun, Mahda M. Bagher, Pejman Sajjadi, and Alexander Klippel. 2020. A Comparison of Visual Attention Guiding Approaches for 360° Image-Based VR Tours. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR).* IEEE, Atlanta, GA, USA, 83–91. https://doi.org/10.1109/VR46266.2020.00026

[34] Mark Weiser and John Seely Brown. 1996. Designing calm technology. *PowerGrid Journal* 1, 1 (1996), 75–85. Publisher: Citeseer.

[35] Glenn F. Wilson. 2002. An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures. *The International Journal of Aviation Psychology* 12, 1 (Jan. 2002), 3–18. https://doi.org/10.1207/S15327108IJAP1201_2

[36] Philipp Wintersberger, Clemens Schartmüller, and Andreas Riener. 2019. Attentive User Interfaces to Improve Multitasking and Take-Over Performance in Automated Driving: The Auto-Net of Things. *International Journal of Mobile Human Computer Interaction* 11, 3 (July 2019), 40–58. https://doi.org/10.4018/IJMHCI.2019070103

[37] Difeng Yu, Ruta Desai, Ting Zhang, Hrvoje Benko, Tanya R. Jonker, and Aakar Gupta. 2022. Optimizing the Timing of Intelligent Suggestion in Virtual Reality. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology.* ACM, Bend OR USA, 1–20. https://doi.org/10.1145/3526113.3545632

A Appendix

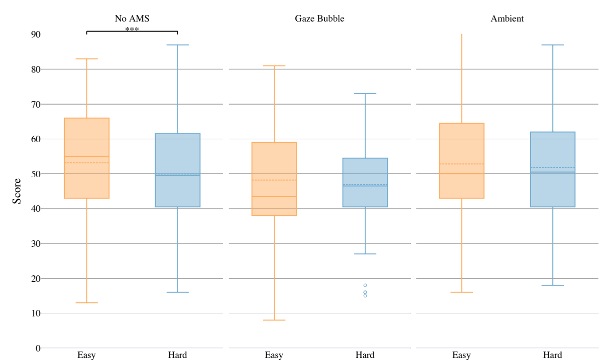


Figure 3: Boxplots of the Score grouped by AMS Design and Difficulty. This plot shows a significant interaction effect between the difficulty of the monitoring task and the AMS design. It shows that while participants perform worse in the difficult condition when they are not assisted by an AMS, their performance in difficult conditions does not decrease when they are assisted by an AMS. Statistically significant differences are indicated as: *p<0.05; **p<0.01; ***p<0.001